

Advancing Interdisciplinary Approaches to Online Safety Research

Senuri Wijenayake

School of Computing Technologies
RMIT University
Melbourne, Victoria, Australia
senuri.wijenayake@rmit.edu.au

Louise La Sala

Orygen, Centre for Youth Mental Health
University of Melbourne
Melbourne, Victoria, Australia
louise.lasala@orygen.org.au

Joanne Gray

Faculty of Arts and Social Sciences
University of Sydney
Sydney, New South Wales, Australia
joanne.gray@sydney.edu.au

Nalin Arachchilage

School of Computing Technologies
RMIT University
Melbourne, Victoria, Australia
nalin.arachchilage@rmit.edu.au

Asangi Jayatilaka

School of Computing Technologies
RMIT University
Melbourne, Victoria, Australia
asangi.jayatilaka@rmit.edu.au

Ryan M. Kelly

School of Computing Technologies
RMIT University
Melbourne, Victoria, Australia
ryan.kelly@rmit.edu.au

Sanchari Das

Department of Information Sciences and Technology
George Mason University
Fairfax, Virginia, USA
sdas35@gmu.edu

Abstract

The growing prevalence of negative experiences in online spaces demands urgent attention from the human-computer interaction (HCI) community. However, research on online safety remains fragmented across different HCI subfields, with limited communication and collaboration between disciplines. This siloed approach risks creating ineffective responses, including design solutions that fail to meet the diverse needs of users, and policy efforts that overlook critical usability concerns. This workshop aims to foster interdisciplinary dialogue on online safety by bringing together researchers from within and beyond HCI – including but not limited to Social Computing, Digital Design, Internet Policy, Cybersecurity, Ethics, and Social Sciences. By uniting researchers, policymakers, industry practitioners, and community advocates we aim to identify shared challenges in online safety research, highlight gaps in current knowledge, and establish common research priorities. The workshop will support the development of interdisciplinary research plans and establish collaborative environments – both within and beyond Australia – to action them.

CCS Concepts

- **Human-centered computing** → *Interaction design; Collaborative and social computing;*
- **Social and professional topics** → *Computing / technology policy;*
- **Security and privacy** → *Human and societal aspects of security and privacy;*

Keywords

online safety, digital harms, interdisciplinary collaboration, inclusive design, user-centred approaches, ethics, policy and governance, industry engagement

ACM Reference Format:

Senuri Wijenayake, Joanne Gray, Asangi Jayatilaka, Louise La Sala, Nalin Arachchilage, Ryan M. Kelly, and Sanchari Das. 2025. Advancing Interdisciplinary Approaches to Online Safety Research. In *37th Australian Conference on Human-Computer Interaction (HCI) (OZCHI '25), November 29–December 03, 2025, Sydney, Australia*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3764687.3767275>

1 Introduction

As online platforms increasingly mediate both our social lives and essential daily tasks – from connecting with others on social media to accessing government services – the issue of user safety has become more urgent than ever [17]. At the same time, negative experiences on online platforms are proliferating at an alarming rate. By 2021, 41% of American adults reported having at least one negative online experience [52], whereas in Australia, the figure was significantly higher – with 70% of adults experiencing at least one negative online interaction in the 12 months leading up to November 2022 [16]. Negative online experiences span a broad spectrum, including but not limited to hate speech and harassment [1, 40, 45], impersonation [43, 58], doxxing [25, 26, 32], misinformation [34, 38, 54], identity-based stereotyping [8, 18, 55–57], non-consensual image sharing [27, 41], and financial scams [11, 43]. These experiences occur across a variety of digital services including social media [3, 48, 48], discussion forums [25, 26, 32], messaging apps [1], financial platforms [43, 58], dating apps [7, 15], and



This work is licensed under a Creative Commons Attribution 4.0 International License.
OZCHI '25, Sydney, Australia
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2016-1/25/11
<https://doi.org/10.1145/3764687.3767275>

online games [12, 23, 33]. Moreover, the rapid adoption of generative AI has further intensified these challenges [24, 28], introducing new threats such as deepfakes [21, 29, 51] and large-scale misinformation campaigns [44, 59] that not only extend existing problems but also create novel ways to target and deceive users.

The current literature on online safety points to a clear conclusion: online safety is a multifaceted and increasingly complex problem that cannot be addressed through any single disciplinary lens. Yet research within HCI remains fragmented, with different sub-fields – including Social Computing, Digital Design, Internet Policy, Cybersecurity, Ethics, and Social Sciences – often tackling specific aspects of the problem in isolation [53]. Through this workshop, we aim to bring together researchers, policymakers, industry practitioners, and community advocates – predominantly within HCI and related disciplines – to critically examine how online safety is conceptualised and investigated across these domains. We will identify current approaches, methods, challenges, and knowledge gaps, and foster interdisciplinary collaboration aimed at advancing more comprehensive, inclusive, effective responses to online safety. A key goal of the workshop is to support the development of a collaborative community of online safety researchers, both within Australia and internationally.

2 Rationale for the Workshop

Online safety has emerged as a significant area of research across multiple sub-fields within HCI, each approaching the issue from different perspectives. For instance, research in CSCW and Social Computing has taken significant efforts to profile online risk experiences, understand users' safety perceptions, and develop targeted interventions to counter online harms [9, 14, 30, 37]. Digital design and UX research has explored the usability and accessibility of safety tools, often highlighting barriers to adoption [2, 4]. Participatory and co-design approaches have centred the experiences of users – especially marginalised communities – in the development of inclusive safety interventions [5, 6, 46, 49].

Moreover, policy work has examined online safety at a systemic level, interrogating platform governance, content moderation practices, regulatory frameworks, and the protection of digital rights [36, 39, 50]. In parallel, cybersecurity research has examined online safety by investigating how users perceive, interpret, and respond to digital threats such as phishing, with a focus on risk awareness, decision-making, and the design of effective training and educational interventions [10, 13, 47].

Furthermore, Critical and Feminist HCI scholars have highlighted how structural inequalities such as gender, race, and ability shape both people's exposure to online harm and the effectiveness of safety interventions [31, 35, 42]. This literature focuses on the perspectives of marginalised users and questions whose needs are prioritised in the design of online platforms. Meanwhile, social sciences has shed light not only on the emotional and behavioural consequences of online harm – such as trauma, coping, and bystander inaction – but also on the psychological drivers of perpetration, victim blaming, and social dynamics that enable or discourage intervention [18–20, 22, 57], all of which are key to designing responsive and effective safety interventions.

Despite this rich and growing body of work, current research efforts remain largely siloed. To move toward more collaborative and impactful approaches to online safety, this workshop has four key objectives, each of which is based on a limitation of the current research landscape.

Objective 1: Build a shared understanding of the current research landscape and promote mutual awareness across different disciplines. There is limited visibility across HCI sub-fields and other related disciplines regarding ongoing research in the online safety domain. Researchers tackling similar problems from different disciplinary perspectives are often unaware of parallel efforts, leading to duplicated work and missed opportunities for collaboration.

Objective 2: Identify research gaps and collaborative opportunities. Research on online safety is frequently conducted in disciplinary silos, making it difficult to recognise shared challenges or determine where interdisciplinary collaboration could generate new insights. A more comprehensive view will help participants engage with broader problem spaces and discover new directions for impactful work.

Objective 3: Initiate the development of an interdisciplinary research agenda that reflects the complexity and urgency of online safety issues. Although many individual projects address important facets of online safety, their scope, methods, and target audiences are often unaligned. A coordinated agenda can help connect these efforts, articulate shared priorities, and direct future research toward high-impact areas.

Objective 4: Establish interdisciplinary networks to support ongoing, multi-stakeholder collaboration. Sustained collaboration between researchers, policymakers, industry practitioners, and community advocates remains rare, despite the complementary expertise each brings. Building these networks is crucial to translating research into real-world impact and implementing the proposed research agenda effectively.

3 Research Areas of Interest

Through structured discussions and collaborative activities, this workshop aims to promote conversation around interdisciplinary approaches to online safety, identifying how HCI research can address key challenges in the field. We invite participation from researchers, policymakers, industry practitioners, and community advocates working on online safety through the lenses of design, policy, technology, ethics, and the social sciences. Contributions are encouraged across the following key areas of online safety research (the full list of topics is provided in Section 7.3):

- **Designing for Safety and Wellbeing:** User-centred and participatory design approaches to create inclusive, effective safety interventions.
- **Technology, AI, and Ethics:** Challenges and innovations in AI-driven safety tools, privacy, transparency, and ethical considerations.
- **Policy, Governance, and Industry:** Regulatory frameworks, cross-platform strategies, and industry perspectives on implementing safety measures.

- **Social, Behavioural, and Cultural Perspectives:** Understanding online harm through social dynamics, cultural contexts, and behavioural interventions.

We particularly encourage submissions that describe interdisciplinary work involving multiple stakeholders across academia, industry, policymakers, and community advocates to address the complex social, technical, policy, and ethical challenges of online safety.

4 Pre-workshop Plans

4.1 Workshop Website

The workshop website will be hosted at www.talkingonlinesafety.org. It will serve as the central hub for all workshop-related information. Prior to the event, we will publish the Call for Participation (see Section 7) on the website, including the workshop objectives, topics of interest, submission instructions, key dates, and organiser profiles. The website will also provide a link to an Expression of Interest (EOI) form for interested participants to complete. The organisers will promote the Call for Participation through professional networks, university mailing lists, research groups, and social media. The organisers will also directly contact potential participants who may be interested, especially those from policy, industry, and community advocacy backgrounds, to ensure diverse representation.

4.2 Expression of Interest (EOI) Form

Those who intend to participate should complete an Expression of Interest (EOI) form ([URL: https://forms.gle/kYF3dwhhRndGENA9A](https://forms.gle/kYF3dwhhRndGENA9A)) by the submission deadline. The EOI form asks participants to provide their name, contact email, affiliation, research or professional area (or sub-field within HCI), accessibility requirements, a brief bio (maximum 300 words) outlining their relevant experience, and an abstract (maximum 300 words) describing their current or recent research, emerging problems, or new perspectives on online safety, as well as what they hope to gain from the workshop.

4.3 Selection Process

The organising committee will review submissions and select participants based on diverse expertise, perspectives, and skills, as well as the quality, novelty, and relevance of their submissions. We aim to select participants to ensure diverse and inclusive representation across career stage, disciplinary background, and stakeholder group, including academia, industry, policy, and community advocacy. We plan to accept 15 - 20 workshop participants. Selected participants will be notified via email by the notification date. Once participants are confirmed, we will post a list of participants and their abstracts on the website to encourage early engagement and visibility.

5 Workshop Structure

This full-day workshop (09:00 - 16:15) is structured in two sessions: a morning session focusing on small group discussions to identify and map participants' research backgrounds, interests, and challenges, and to explore opportunities for interdisciplinary, multi-stakeholder research into online safety. The afternoon session is dedicated to

collaborative brainstorming of potential research proposals that participants can develop and continue exploring after the workshop.

5.1 Inclusion and Accessibility Considerations

The organising committee is committed to fostering an inclusive and accessible workshop environment. We will adhere to the conference's accessibility and inclusion guidelines in all aspects of planning and delivery. A wheelchair-accessible venue has been requested, and participants will be invited to indicate any accessibility requirements in advance via the Expression of Interest form so that appropriate arrangements can be made. Workshop materials will be provided in accessible digital or printed formats, depending on participant preference. Activities are designed to support a range of participation styles – including verbal, written, and visual contributions – to accommodate neurodiverse participants and those with varied communication preferences.

5.2 Morning Session

9:00 - 9:30 Workshop Opening & Participant Profiles: The session will begin with introductions from the workshop organisers, followed by an overview of the workshop objectives and agenda. We expect around 15–20 participants from fields such as HCI, Digital Design, Internet Policy, Cybersecurity, Ethics, and the Social Sciences, alongside policymakers, community advocates, and industry representatives. The workshop organisers are either from these disciplines or have established connections with these stakeholders and will leverage their networks to encourage participation, ensuring diverse perspectives.

Next, participants will create brief research profiles using a digital template provided by the organisers. These profiles will include their name, affiliation, contact information, research interests, methodological expertise, and any collaboration or funding opportunities they are exploring. The completed profiles will be accessible to participants during the workshop via a shared folder and will be distributed by email and posted on the workshop website after the event to support ongoing networking and collaboration.

9:30 - 10:45 Mapping Participants' Research Interests: Participants will be organised into interdisciplinary groups of 4–5 members, based on research interests identified through the pre-workshop EOI form. Each group member will present their research through a structured introduction covering their current work on online safety and its contribution to the field, the key research questions they are investigating, current challenges they face, and future opportunities they see for research. Participants may refer to their profiles during these presentations.

Following the introductions, the groups will engage in structured discussions using the following prompts:

- How does each member's research address online safety concerns?
- What research questions are emerging from their different perspectives?
- What barriers are limiting progress in their areas?
- What unexplored opportunities exist for collaborative research?

During these discussions, groups will collaboratively map their research landscape using sticky notes to visualise shared research questions, different approaches, and common challenges across disciplines.

10:45 – 11:00 Morning Tea Break

11:00 – 12:00 Group Presentations: Each group will present their discussion findings in 5-minute presentations. Collectively, these presentations will provide an overview of the current research landscape, highlighting different disciplinary approaches and common challenges. During the presentations, workshop organisers will facilitate discussions around similarities and differences between groups, questions about various approaches, and potential areas for collaboration. This process will help all participants understand the broader research context and set the foundation for collaborative idea development in the afternoon session.

12:00 – 13:00 Lunch Break

5.3 Afternoon Session

13:00 – 14:00 Developing a Mini-Project Proposal: After lunch, participants will remain in their morning groups to develop a project proposal addressing a specific research gap or challenge identified during earlier discussions. Proposals may draw on insights from the group presentations or broader workshop conversations. Each proposal should outline the project aims, proposed methodology, expected outcomes, and how each member's expertise will contribute. Groups will create a slide deck of up to 3 slides to summarise their proposal.

The goal of this activity is to develop feasible, interdisciplinary research projects that participants are genuinely interested in pursuing. These proposals will serve as a foundation for ongoing collaboration beyond the workshop. After the event, summaries of all proposals and discussion outcomes will be circulated to attendees via email and posted on the workshop website.

14:00 - 14:45 Project Presentations: Groups will present their research proposals in 10-minute presentations to share innovative project ideas and receive constructive feedback from all workshop participants. Participants will be encouraged to provide feedback, suggest improvements, and identify potential connections between different project proposals.

14:45 - 15:00 Afternoon Tea Break

15:00 – 16:00 Expert Panel Discussion: A panel of online safety experts will reflect on the day's discussions to provide external perspective on the research directions discussed throughout the workshop. The panel will include 4-5 members representing academics, policy experts, community advocates, and industry representatives, comprising both organising committee members and invited guests recruited through the committee's networks. A workshop organiser will moderate the panel discussion for 45 minutes, followed by a 15-minute question and answer session where workshop participants can seek advice and explore potential partnerships with the expert panel members.

16:00 – 16:15 Closing and Next Steps: The workshop will conclude with a summary of key outcomes and identified collaborations, along with information on follow-up activities. The aim is to ensure participants leave with clear pathways to continue the collaborations and research directions developed during the workshop.

6 Post-workshop Publication Plan

After the workshop, organisers will compile the researcher profiles and project proposals into a digital format and share them with attendees via email and post on the workshop website. This resource is expected to facilitate ongoing collaboration among participants and attract other researchers interested in online safety to explore partnership opportunities with workshop attendees.

7 Call for Participation: Advancing Interdisciplinary Approaches to Online Safety Research

- Sydney, Australia
- Workshop date: 30 November 2025
- Website: www.talkingonlinesafety.org
- Contact: Senuri Wijenayake [senuri.wijenayake@rmit.edu.au]

7.1 Important Dates

- Call for Participation: 19 September 2025
- Submission Deadline: 19 October 2025 23:59 (AoE)
- Notification of Acceptance: 31 October 2025 23:59 (AoE)
- Workshop Dates: 30 November 2025 09:00 - 16:15, Sydney (Australia)

7.2 Workshop Overview

Online safety is a critical and complex challenge that affects users across diverse digital platforms. Despite substantial research efforts within Human-Computer Interaction (HCI) and related fields, work remains fragmented across disciplines such as Social Computing, Design, Policy, Cybersecurity, Ethics, and Social Sciences. This workshop aims to unite researchers, policymakers, industry practitioners, and community advocates to bridge these silos. Together, participants will share knowledge, identify gaps, and foster interdisciplinary collaboration to develop comprehensive, inclusive, and impactful solutions addressing the multifaceted nature of online safety.

The workshop will facilitate dialogue to:

- Build a shared understanding of the current research landscape.
- Identify research gaps and opportunities for collaboration.
- Develop coordinated interdisciplinary research agendas.
- Establish sustained networks for multi-stakeholder collaboration.

7.3 Topics of Interest

Topics of interest include, but are not limited to:

- **Designing for Safety and Wellbeing**
 - User-centred approaches to understanding and responding to online abuse

- *Safety by design* principles and their practical implementation across platforms
- Designing for resilience and digital literacy in online safety
- Participatory design methods for developing safety interventions with affected communities
- Prevention-focused approaches to online safety
- Evaluation methods for assessing safety intervention effectiveness
- Psychological impacts of online harm and trauma-informed design
- Cultural and contextual considerations in online safety design, particularly for marginalised groups
- **Technology, AI, and Ethics**
 - Emerging online safety challenges from generative AI and synthetic media
 - Use of AI for enhancing online safety and harm detection
 - Transparency and explainability in safety-related AI systems
 - Privacy-preserving safety interventions
 - Ethical considerations in developing safety technologies, especially for marginalised or vulnerable users
 - Evaluating the impact and unintended consequences of AI-driven safety tools
- **Policy, Governance, and Industry**
 - Policy frameworks and regulatory approaches to online safety
 - Inclusion of equity-focused safety standards and obligations for protecting marginalised users
 - Industry perspectives on implementing and scaling safety features
 - Cross-platform safety considerations and interoperability
 - Collaborative models involving academia, industry, policymakers, and communities
- **Social, Behavioural, and Cultural Perspectives**
 - Social and behavioural dynamics of online abuse (perpetrator, bystander, and survivor perspectives)
 - Cross-cultural differences in perceptions of safety, harm, and acceptable behaviour online
 - Structural, cultural, and socioeconomic factors influencing online safety for underrepresented communities
 - Psychological impacts of sustained exposure to online harm and strategies for resilience
 - Online safety education, awareness, and digital literacy interventions from a behavioural perspective
 - Collective action, social movements, and advocacy around online safety and digital rights

We particularly encourage submissions that describe interdisciplinary work involving multiple stakeholders across academia, industry, policymakers, and community advocates to address the complex social, technical, policy, and ethical challenges of online safety.

7.4 Who Should Participate?

We welcome submissions from researchers and practitioners across diverse disciplines such as HCI, Social Computing, Design, Policy, Cybersecurity, Ethics, and Social Sciences. Policymakers, industry professionals, and community advocates engaged in digital safety are encouraged to join. Both emerging and established researchers working on new or ongoing online safety projects will find this workshop valuable.

7.5 How to Participate

To participate, please submit an Expression of Interest (EOI) using this form: <https://forms.gle/kYF3dwhhRndGENA9A>. Your submission should include an abstract of up to 300 words describing your current or recent research, theoretical discussions, emerging challenges, or new perspectives related to online safety. Contributions from both early-stage and completed work are welcome. Selected participants will be invited to join a collaborative, interactive workshop focused on co-developing research agendas and building lasting interdisciplinary networks.

7.6 Organisers

Senuri Wijenayake is a Lecturer in the School of Computing Technologies at RMIT University. Her interdisciplinary research spans Human-Computer Interaction (HCI), Social Psychology, and Design to advance online safety for marginalised communities, including women and gender-diverse users. She examines the unique harms these communities experience, critiques the limitations of current technological responses, and uses participatory design methods to co-develop tailored safety interventions. Senuri was recently awarded a research grant from the Australian Communications Consumer Action Network (ACCAN) to develop design strategies and policy recommendations addressing technology-facilitated abuse on social media, in collaboration with WESNET, the eSafety Commissioner, and the Victorian Pride Centre.

Joanne Gray is a Lecturer in Digital Cultures in the Discipline of Media and Communications, Faculty of Arts and Social Sciences. She is an interdisciplinary academic with expertise in digital platform policy and governance. Her research seeks to understand how digital platforms, such as Google/Alphabet and Facebook/Meta, exercise private power and explore relevant policy options. Dr Gray is also the Commissioning Editor for the journal *Policy & Internet*.

Asangi Jayathilaka is a Lecturer in Cybersecurity and Software Systems at RMIT University, Melbourne. She brings extensive experience in interdisciplinary research and participatory design, including co-design, specifically focused on empowering marginalised communities to shape their digital technology experiences. Her work deeply engages with diverse groups, including individuals with cognitive impairments, women, gender-diverse people, and culturally and linguistically diverse users.

Louise La Sala is a Research Fellow at Orygen, Centre for Youth Mental Health at the University of Melbourne. Her research is focused on youth self-harm and suicide prevention, with a specific interest in the impact of social media on the mental health and well-being of young people. Dr La Sala is a lead researcher on the #chatsafe program of work which aims to provide young people,

educators, and families with the tools to communicate safely online about self-harm and suicide. Her work investigates the complex relationship between social media and youth mental health, and she brings unique expertise in developing effective strategies to promote online safety and prevent self-harm and suicide among young people. Her work has informed online safety and suicide prevention policy at both a national and international level and she regularly works with popular social media platforms.

Nalin Arachchilage is an Associate Professor in the School of Computing Technologies at RMIT University. His research focuses on usable security and privacy, particularly at the intersection of computer security and human-computer interaction. He has developed innovative approaches to online safety, including a game design framework to educate users on protecting themselves from phishing attacks. Nalin's work bridges cybersecurity, HCI, and software engineering to create effective, user-centred security solutions.

Ryan M. Kelly is an Associate Professor in the School of Computing Technologies at RMIT University. He is interested in designing to enable safe experiences for older adults online, particularly in the areas of social interaction, digital health and finance. His recent research has focused on designing for safe video calling experiences between older adults and conversational volunteers. Ryan's disciplinary background is in Applied Psychology, HCI and CSCW.

Sanchari Das is an Assistant Professor in the Department of Information Sciences and Technology at George Mason University, USA. Her research focuses on usable security and privacy, and applied AI, particularly in domains such as IoT, AR/VR/MR, healthcare, finance, education, and everyday digital interactions. She directs the CAPS (Center for AI, Privacy, and Security) Lab and co-directs the Secure Realities Lab, where she brings an interdisciplinary approach combining AI/ML, human-computer interaction, and cybersecurity to design resilient socio-technical systems. Prior to academia, she worked in industry as a Security and Software Engineer at American Express, Infosys, and HCL Technologies, and also served as a User Experience Consultant and Global Privacy Adviser for various organizations.

Acknowledgments

This work was supported by the Australian Communications Consumer Action Network (ACCAN). The operation of ACCAN is made possible by funding provided by the Commonwealth of Australia under section 593 of the Telecommunications Act 1997. This funding is recovered from charges on telecommunications carriers.

References

- [1] Dhruv Agarwal, Farhana Shahid, and Aditya Vashistha. 2024. Conversational Agents to Facilitate Deliberation on Harmful Content in WhatsApp Groups. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (Nov. 2024), 1–32. doi:10.1145/3687030 Publisher: Association for Computing Machinery (ACM).
- [2] Zainab Agha. 2023. To Nudge or Not to Nudge: Co-Designing and Evaluating the Effectiveness of Adolescent Online Safety Nudges. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (Chicago, IL, USA) (IDC '23). Association for Computing Machinery, New York, NY, USA, 760–763. doi:10.1145/3585088.3593923
- [3] Zainab Agha, Karla Badillo-Urquiola, and Pamela J. Wisniewski. 2023. "Strike at the Root": Co-designing Real-Time Social Media Interventions for Adolescent Online Risk Prevention. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 1–32. doi:10.1145/3579625 Publisher: Association for Computing Machinery (ACM).
- [4] Zainab Agha, Jinkyung Park, Ruyuan Wan, Naima Samreen Ali, Yiwei Wang, Dominic Difranzo, Karla Badillo-Urquiola, and Pamela J. Wisniewski. 2024. Tricky vs. Transparent: Towards an Ecologically Valid and Safe Approach for Evaluating Online Safety Nudges for Teens. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 116, 20 pages. doi:10.1145/3613904.3642313
- [5] Zainab Agha, Zinan Zhang, Oluwatomisin Obajemu, Luke Shirley, and Pamela J. Wisniewski. 2022. A Case Study on User Experience Bootcamps with Teens to Co-Design Real-Time Online Safety Interventions. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (CHI EA '22). Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/3491101.3503563
- [6] Sadiq Aliyu, Sushmita Khan, Aminata N. Mbodi, Oluwafemi Osho, Lingyuan Li, Bart Knijnenburg, and Mauro Cherubini. 2024. Participatory Design to Address Disclosure-Based Cyberbullying. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) (DIS '24). Association for Computing Machinery, New York, NY, USA, 1547–1565. doi:10.1145/3643834.3660716
- [7] Hanan Khalid Aljasim and Douglas Zytko. 2023. Foregrounding Women's Safety in Mobile Social Matching and Dating Apps: A Participatory Design Study. *Proceedings of the ACM on Human-Computer Interaction* 7, GROUP (Jan. 2023), 1–25. doi:10.1145/3567559 Publisher: Association for Computing Machinery (ACM).
- [8] Rula Odeh Alsawala and Maisaa N. Alrawashdeh. 2022. The role of patriarchal structure and gender stereotypes in cyber dating abuse: A qualitative examination of male perpetrators experiences. *The British Journal of Sociology* 73, 3 (June 2022), 587–606. doi:10.1111/1468-4446.12946
- [9] Ashwaq Alsobai, Afsaneh Razai, Zainab Agha, Shiza Ali, Gianluca Stringhini, Munmun De Choudhury, and Pamela J. Wisniewski. 2024. Profiling the Offline and Online Risk Experiences of Youth to Develop Targeted Interventions for Online Safety. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 114:1–114:37. doi:10.1145/3637391
- [10] Nalin Asanka Gamagedara Arachchilage and Steve Love. 2014. Security awareness of computer users: A phishing threat avoidance perspective. *Computers in human behavior* 38 (2014), 304–312. https://www.sciencedirect.com/science/article/pii/S0747563214003331 Publisher: Elsevier.
- [11] Rosanna Bellini. 2023. Paying the Price: When Intimate Partners Use Technology for Financial Harm. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–17. doi:10.1145/3544548.3581101
- [12] Nicole A Beres, Julian Frommel, Elizabeth Reid, Regan L Mandryk, and Madison Klarkowski. 2021. Don't You Know That You're Toxic: Normalization of Toxicity in Online Gaming. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–15. doi:10.1145/3411764.3445157
- [13] Nathan Beau, Asangi Jayatilaka, Manssoreh Zahedi, Muhammad Ali Babar, Laura Hartley, Winston Lewinsmith, and Irina Baetu. 2023. Falling for phishing attempts: An investigation of individual differences that are associated with behavior in a naturalistic phishing simulation. *Computers & Security* 131 (2023), 103313. https://www.sciencedirect.com/science/article/pii/S0167404823002237 Publisher: Elsevier.
- [14] Youjin Choe, Senuri Wijenayake, Martin Tomko, and Mohsen Kalantari. 2024. Mapping in harmony: Co-designing user interfaces for conflict management on OSM. *International Journal of Human-Computer Studies* 190 (2024), 103316. https://www.sciencedirect.com/science/article/pii/S1071581924001009?casa_token=0eVq2KoQyKkAAAAA:lWmcNuJq15KFqfTtba6hl_KRsk_TmRBqBhurTIV4OX22l8tqgEd0BR5Pvv43FopNwaoESyRbnYE Publisher: Elsevier.
- [15] Isha Datey and Douglas Zytko. 2024. "Just Like, Risking Your Life Here": Participatory Design of User Interactions with Risk Detection AI to Prevent Online-to-Offline Harm Through Dating Apps. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (Nov. 2024), 1–41. doi:10.1145/3686906 Publisher: Association for Computing Machinery (ACM).
- [16] eSafety Commissioner. 2022. Australians' negative online experiences 2022. https://www.esafety.gov.au/research/australians-negative-online-experiences-2022
- [17] eSafety Commissioner. 2025. Encounters with online hate. https://www.esafety.gov.au/research/encounters-with-online-hate
- [18] Diane Felmlee, Paulina Inara Rodis, and Amy Zhang. 2020. Sexist Slurs: Reinforcing Feminine Stereotypes Online. *Sex Roles* 83, 1-2 (July 2020), 16–28. doi:10.1007/s11199-019-01095-z
- [19] Asher Flynn, Elena Cama, and Adrian J. Scott. 2025. Attitudes Towards Image-Based Sexual Abuse. In *Palgrave Studies in Cybercrime and Cybersecurity*. Springer Nature Switzerland, Cham, 49–69. doi:10.1007/978-3-031-83647-3_3 ISSN: 2946-2770, 2946-2789.
- [20] Asher Flynn, Anastasia Powell, and Sophie Hindes. 2024. An intersectional analysis of technology-facilitated abuse: Prevalence, experiences and impacts of victimization. *The British Journal of Criminology* 64, 3 (2024), 600–619. https://academic.oup.com/bjc/article-abstract/64/3/600/7259611 Publisher: Oxford University Press UK.

- [21] Asher Flynn, Anastasia Powell, Adrian J. Scott, and Elena Cama. 2022. Deepfakes and digitally altered imagery abuse: A cross-country exploration of an emerging form of image-based sexual abuse. *The British Journal of Criminology* 62, 6 (2022), 1341–1358. <https://academic.oup.com/bjc/article-abstract/62/6/1341/6448791> Publisher: Oxford University Press UK.
- [22] Asher Flynn, Anastasia Powell, and Lisa Wheildon. 2024. *Workplace technology-facilitated sexual harassment: perpetration, responses and prevention*. Australia's National Research Organisation for Women's Safety (ANROWS), Australia. <https://apo.org.au/node/326578>
- [23] Guo Freeman, Julian Frommel, Regan L Mandryk, Jan Gugenheimer, Lingyuan Li, and Daniel Johnson. 2024. Novel Approaches for Understanding and Mitigating Emerging New Harms in Immersive and Embodied Virtual Spaces: A Workshop at CHI 2024. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–7. doi:10.1145/3613905.3636288
- [24] Guo Freeman, Douglas Zytko, Afshaneh Razi, Cliff Lampe, Heloisa Candello, Timo Jakobi, and Konstantin Kosta Aal. 2025. New Opportunities, Risks, and Harm of Generative AI for Fostering Safe Online Communities. In *Companion Proceedings of the 2025 ACM International Conference on Supporting Group Work (GROUP '25)*. Association for Computing Machinery, New York, NY, USA, 2–5. doi:10.1145/3688828.3700747
- [25] Nitesh Goyal, Leslie Park, and Lucy Vasserman. 2022. "You have to prove the threat is real": Understanding the needs of Female Journalists and Activists to Document and Report Online Harassment. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–17. doi:10.1145/3491102.3517517
- [26] Catherine Han, Anne Li, Deepak Kumar, and Zakir Durumeric. 2024. PressProtect: Helping Journalists Navigate Social Media in the Face of Online Harassment. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (Nov. 2024), 1–34. doi:10.1145/3687048 Publisher: Association for Computing Machinery (ACM).
- [27] Rakibul Hasan, Bennett I. Bertenthal, Kurt Hugenberg, and Apu Kapadia. 2021. Your Photo is so Funny that I don't Mind Violating Your Privacy by Sharing it: Effects of Individual Humor Styles on Online Photo-sharing Behaviors. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. doi:10.1145/3411764.3445258
- [28] Will Hawkins, Brent Mittelstadt, and Chris Russell. 2025. Deepfakes on Demand: The rise of accessible non-consensual deepfake image generators. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 1602–1614. doi:10.1145/3715275.3732107
- [29] Nicola Henry and Alice Witt. 2021. *Governing Image-Based Sexual Abuse: Digital Platform Policies, Tools, and Practices*. Emerald Publishing Limited. 749–768 pages. doi:10.1108/978-1-83982-848-520211054
- [30] Naulsberry Jean Baptiste, Jinkyung Park, Neeraj Chatlani, Naima Samreen Ali, and Pamela J. Wisniewski. 2023. Teens on Tech: Using an Asynchronous Remote Community to Explore Adolescents' Online Safety Perspectives. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) (CSCW '23 Companion). Association for Computing Machinery, New York, NY, USA, 45–49. doi:10.1145/3584931.3606967
- [31] Dahlia Jovic, Ren Galwey, Lucy A. Sparrow, Mahli-Ann Butt, Yige Song, Sable Wang-Wills, and Eduardo A. Oliveira. 2025. The AI Ally Project: Designing a Human-Centred AI Tool for Women's Safety in Online Spaces. In *Companion Proceedings of the ACM on Web Conference 2025* (Sydney NSW, Australia) (WWW '25). Association for Computing Machinery, New York, NY, USA, 2791–2795. doi:10.1145/3701716.3716878
- [32] Song Mi Lee, Andrea K. Thomer, and Cliff Lampe. 2022. The Use of Negative Interface Cues to Change Perceptions of Online Retributive Harassment. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–23. doi:10.1145/3555226 Publisher: Association for Computing Machinery (ACM).
- [33] Daniel Maddem, Yuxuan Liu, Haowei Yu, Mustafa Feyyaz Sonbudak, Giovanni M Troiano, and Casper Harteveld. 2021. "Why Are You Playing Games? You Are a Girl!": Exploring Gender Biases in Esports. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–15. doi:10.1145/3411764.3445248
- [34] Lisa Mekioussa Malki, Dilisha Patel, and Aneesha Singh. 2024. "The Headline Was So Wild That I Had To Check": An Exploration of Women's Encounters With Health Misinformation on Social Media. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (April 2024), 1–26. doi:10.1145/3637405 Publisher: Association for Computing Machinery (ACM).
- [35] Tara Matthews, Eli Bursztein, Patrick Gage Kelley, Lea Kissner, Andreas Kramm, Andrew Oplinger, Andreas Schou, Manya Sleeper, Stephan Somogyi, Dalila Szostak, Kurt Thomas, Anna Turner, Jill Palzkill Woelfer, Lawrence L. You, Izzie Zahorian, and Sunny Consolvo. 2025. Supporting the Digital Safety of At-Risk Users: Lessons Learned from 9+ Years of Research and Training. *ACM Trans. Comput.-Hum. Interact.* 32, 3 (June 2025), 22:1–22:39. doi:10.1145/3716382
- [36] Rachel Elizabeth Moran, Joseph Schafer, Mert Bayar, and Kate Starbird. 2025. The End of Trust and Safety?: Examining the Future of Content Moderation and Upheavals in Professional Online Safety Efforts. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 176, 14 pages. doi:10.1145/3706598.3713662
- [37] Oluwatomisin Obajemu, Zainab Agha, Farzana A. Chowdhury, and Pamela J. Wisniewski. 2024. Towards Enforcing Good Digital Citizenship: Identifying Opportunities for Adolescent Online Safety Nudges. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 136:1–136:37. doi:10.1145/3637413
- [38] Wei Peng, Hee Rin Lee, and Sue Lim. 2024. Leveraging chatbots to combat health misinformation for older adults: Participatory design study. *JMIR Formative Research* 8, 1 (2024), e60712. <https://formative.jmir.org/2024/1/e60712/> Publisher: JMIR Publications Inc., Toronto, Canada.
- [39] Andy Phippen. 2025. Online Safety Policy—Moving On? In *Policy and Rights Challenges in Children's Online Behaviour and Safety, 2017–2023*, Andy Phippen (Ed.). Springer Nature Switzerland, Cham, 1–18. doi:10.1007/978-3-031-80286-7_1
- [40] Kaike Ping, James Hawdon, and Eugenia H Rho. 2025. Perceiving and Countering Hate: The Role of Identity in Online Responses. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (May 2025), 1–28. doi:10.1145/3711045 Publisher: Association for Computing Machinery (ACM).
- [41] Li Qiwei, Allison McDonald, Oliver L. Haimson, Sarita Schoenebeck, and Eric Gilbert. 2024. The Sociotechnical Stack: Opportunities for Social Computing Research in Non-Consensual Intimate Media. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (Nov. 2024), 1–21. doi:10.1145/3686914 Publisher: Association for Computing Machinery (ACM).
- [42] Casey Randazzo, Carol F. Scott, Rosanna Bellini, Tawfiq Ammari, Michael Ann Devito, Bryan Semaan, and Nazanin Andalibi. 2023. Trauma-Informed Design: A Collaborative Approach to Building Safer Online Spaces. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) (CSCW '23 Companion). Association for Computing Machinery, New York, NY, USA, 470–475. doi:10.1145/3584931.3611277
- [43] Lubna Razzaq, Tallal Ahmad, Samia Ibtasam, Umer Ramzan, and Shrirang Mare. 2021. "We Even Borrowed Money From Our Neighbor": Understanding Mobile-based Frauds Through Victims' Experiences. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–30. doi:10.1145/3449115 Publisher: Association for Computing Machinery (ACM).
- [44] Christian Reuter, Amanda Lee Hughes, and Cody Buntain. 2024. Combating information warfare: state and trends in user-centred countermeasures against fake news and misinformation. *Behaviour & Information Technology* 44, 13 (Dec. 2024), 1–14. doi:10.1080/0144929x.2024.2442486 Publisher: Informa UK Limited.
- [45] Hyeyoung Ryu and Sungka Kang. 2025. Exploring Design Tensions in Countering Microaggressions in Online Communities for Stigmatized Health Support. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–9. doi:10.1145/3706599.3706702
- [46] Louise La Sala, Amanda Sabo, Maria Michail, Pinar Thorn, Michelle Lamblin, Vivienne Browne, and Jo Robinson. 2025. Online Safety When Considering Self-Harm and Suicide-Related Content: Qualitative Focus Group Study With Young People, Policy Makers, and Social Media Industry Professionals. *Journal of Medical Internet Research* 27, 1 (March 2025), e66321. doi:10.2196/66321
- [47] Orville Sarker, Asangi Jayatilaka, Sherif Haggag, Chelsea Liu, and M. Ali Babar. 2024. A Multi-vocal Literature Review on challenges and critical success factors of phishing education, training and awareness. *Journal of Systems and Software* 208 (2024), 111899. <https://www.sciencedirect.com/science/article/pii/S0164121223002947> Publisher: Elsevier.
- [48] Kelsea Schulenberg, Guo Freeman, Lingyuan Li, and Catherine Barwulor. 2023. "Creepy Towards My Avatar Body, Creepy Towards My Body": How Women Experience and Manage Harassment Risks in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (Sept. 2023), 1–29. doi:10.1145/3610027
- [49] Sharanya Shanmugam and Mark Findlay. 2024. Empowering Young Digital Citizens: The Call for Co-creation Between Youth and Policymakers in Regulating for Online Safety and Privacy. In *Mobile Media Use Among Children and Youth in Asia*, Andrew Zi Han Yee (Ed.). Springer Netherlands, Dordrecht, 159–184. doi:10.1007/978-94-024-2282-5_9
- [50] Markus Trengove, Emre Kazim, Denise Almeida, Airlie Hilliard, Sara Zannone, and Elizabeth Lomas. 2022. A critical review of the Online Safety Bill. *Patterns* 3, 8 (Aug. 2022). doi:10.1016/j.patter.2022.100544 Publisher: Elsevier.
- [51] Rebecca Umbach, Nicola Henry, Gemma Faye Beard, and Colleen M. Berryessa. 2024. Non-Consensual Synthetic Intimate Imagery: Prevalence, Attitudes, and Knowledge in 10 Countries. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–20. doi:10.1145/3613904.3642382
- [52] Emily A. Vogels. 2021. The State of Online Harassment. <https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/>
- [53] Ashley Marie Walker, Michael Ann DeVito, Karla Badillo-Urquilo, Rosanna Bellini, Stevie Chancellor, Jessica L. Feuston, Kathryn Henne, Patrick Gage Kelley, Shalaleh Rismani, Renee Shelby, and Renwen Zhang. 2024. "What is Safety?": Building Bridges Across Approaches to Digital Risks and Harms. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing* (Minneapolis, MN, USA) (CSCW '24 Companion). Association for Computing Machinery, New York, NY, USA, 470–475. doi:10.1145/3584931.3611277

- Social Computing*. ACM, San Jose Costa Rica, 736–739. doi:10.1145/3678884.3681824
- [54] Senuri Wijenayake, Danula Hettichchi, Simo Hosio, Vassilis Kostakos, and Jorge Goncalves. 2020. Effect of conformity on perceived trustworthiness of news in social media. *IEEE Internet Computing* 25, 1 (2020), 12–19. doi:10.1109/MIC.2020.3032410 Publisher: IEEE.
- [55] Senuri Wijenayake, Jolan Hu, Vassilis Kostakos, and Jorge Goncalves. 2021. Quantifying the Effects of Age-Related Stereotypes on Online Social Conformity. In *Human-Computer Interaction – INTERACT 2021*, Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen (Eds.). Vol. 12935. Springer International Publishing, Cham, 451–475. doi:10.1007/978-3-030-85610-6_26 Series Title: Lecture Notes in Computer Science.
- [56] Senuri Wijenayake, Niels Van Berkela, Vassilis Kostakos, and Jorge Goncalves. 2019. Measuring the Effects of Gender on Online Social Conformity. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–24. doi:10.1145/3359247
- [57] Michelle F. Wright and Sebastian Wachs. 2020. Adolescents' cyber victimization: The influence of technologies, gender, and gender stereotype traits. *International journal of environmental research and public health* 17, 4 (2020), 1293. https://www.mdpi.com/1660-4601/17/4/1293 Publisher: MDPI.
- [58] Yuxiang Zhai, Xiao Xue, Zekai Guo, Tongtong Jin, Yuting Diao, and Jihong Jeung. 2025. Hear Us, then Protect Us: Navigating Deepfake Scams and Safeguard Interventions with Older Adults through Participatory Design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–19. doi:10.1145/3706598.3714423
- [59] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–20. doi:10.1145/3544548.3581318